Available online at www.sciencedirect.com

ScienceDirect

Energy Procedia 111 (2017) 1070 – 1078

Energy Procedia

8th International Conference on Sustainability in Energy and Buildings, SEB-16, 11-13 September 2016, Turin, ITALY

Assessment of building operational performance using data mining techniques: a case study

Cheng Fan, Fu Xiao*

Department of Building Services Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

Abstract

Today's buildings are not only energy intensive, but also information intensive. Massive amounts of operational data are available for knowledge discovery. Data mining (DM) has excellent ability in extracting insights from massive data. This paper performs a case study on the assessment of building operational performance using DM techniques. Typical DM techniques are compared and considerations for choosing specific DM techniques for the case study are presented. The methodology developed has been applied to analyze the data retrieved from a university building in Hong Kong. Useful insights have been obtained to identify typical operation patterns and energy conservation opportunities.

Peer-review under responsibility of KES International.

Keywords: Building Energy Conservation; Building Automation System; Data Mining; Big Data; Intelligent Building.

1. Introduction

Buildings have become one of the largest energy consumers around the world. The energy saving potential in building operations is huge due to the widespread occurrence of equipment degradation, faults in system components, and deficiencies in control strategies in buildings. Advanced technologies, such as the building automation system (BAS), have been integrated with modern buildings to facilitate the real-time monitoring and controls over building operations. Massive amounts of building operational data are collected and stored in BAS, from which valuable insights can be extracted to enhance the building operational performance. Nevertheless, building data are far from being fully utilized, mainly due to the lack of methods and tools for handling those big

* Corresponding author. Tel.: 852-27664194; fax: 852-27657198
E-mail address: linda.xiao@polyu.edu.hk

Peer-review under responsibility of KES International.
doi:10.1016/j.egypro.2017.03.270

data. Conventional methods for using the building data, which primarily rely on physical principles, statistics and engineering expertise, are neither efficient nor effective in discovering potentially useful yet previously unknown knowledge from massive BAS data sets. Advanced methodologies and tools are urgently needed in the building field to tackle the big data challenge.

Data mining (DM) technology is a promising solution and renowned for its excellent ability in extracting useful insights from massive data sets. It has been widely used in various industries, including the financial services, retails, health care, and even counter-terrorism [1, 2]. In general, DM techniques can be classified into two groups, i.e., supervised and unsupervised learning. Supervised learning techniques aims to perform regression or classification based on the relationships discovered between input and output variables. The knowledge discovered is usually represented using various models. Supervised learning techniques have been widely applied for energy consumption prediction [3-6] and fault detection and diagnosis [7-10] in the building field. One intrinsic limitation of supervised learning is that it needs reliable training data, which are very hard to obtain in building operations, particularly data under fault conditions. By contrast, unsupervised learning doesn't have such a need and it focuses on discovering the intrinsic structures, correlations and associations in the data. Moreover, it requires less domain expertise which makes it more preferable in real applications to discover new knowledge. The knowledge obtained using unsupervised DM techniques is usually in the form of data clusters, association rules, or anomalies.

This paper performs a case study on extracting useful knowledge from massive building operational data using DM techniques and their potential applications in building energy management. The methodology is derived from the generic data analytic framework, which was proposed in our previous study [11]. The main DM techniques adopted are decision trees and association rule mining. The methodology has been applied to analyze the data retrieved from one building in the Hong Kong Polytechnic University. The results show that useful insights can be obtained for enhancing building energy efficiency.

2. Research Methodology

2.1. Research outline

The knowledge gap between building professionals and advanced analytics motivated us to develop a generic DM-based analytic framework for analyzing big building operational data. Based on extensive investigation of popular DM techniques and deep understanding of building operations, a framework has been proposed in our previous paper [11]. The framework contains 4 phases, i.e., data exploration, data partitioning, knowledge discovery and post-mining. The data exploration phase mainly aims to enhance the data quality and prepares the data into suitable formats for the following data analysis. The data partitioning phase intends to improve the reliability and sensitivity of the knowledge discovered by dividing the building operational data into several groups according to the characteristics of building operations. Various DM techniques can be adopted to extract knowledge at the knowledge discovery phase. Domain expertise is involved in the post-mining phase to interpret, select and apply potentially useful knowledge.

The methodology adopted in this paper is derived from the framework. The clustering analysis method and the decision tree method are compared and the latter is chosen for data partitioning. The quantitative association rule mining is applied for knowledge discovery. The details are introduced in the following sub-sections.

2.2. Data partitioning

Building operations are highly complicated due to the constantly changing indoor and outdoor conditions. It is therefore not wise to treat the building operational data as a whole for data analysis, as it will negatively affect the reliability and sensitivity of knowledge discovered. Typical building operational data are stored in a two-dimensional data table, in which each column represents a variable and each row stores the values of different variables sampled at the same instant of time. Data partitioning refers to the process of dividing the entire data table into several subsets, each containing a number of rows.

Two types of methods are suitable for partitioning building operational data. The first is to treat each row as an observation and then grouping observations based on their similarities which can be evaluated by Euclidean distance